UCLA

Title
Motion Integration Using Competitive Priors

Permalink
https://escholarship.org/uc/item/7w9500qn

Authors
Shuang Wu
Hongjing Lu
Alan Lee
et al.

Publication Date
2011-10-25

Motion integration using competitive priors

Shuang Wu[1], Hongjing Lu[2], Alan Lee[2], and Alan Yuille[1]

[1] Department of Statistics, UCLA
[2] Department of Psychology, UCLA

Abstract. Psychophysical experiments show that humans are better at perceiving rotation and expansion than translation [4][8]. These findings are inconsistent with standard models of motion integration which predict best performance for translation. To explain this discrepancy, our theory formulates motion perception at two levels of inference: we first perform model selection between the competing models (e.g. translation, rotation, and expansion) and then estimate the velocity using the selected model. We define novel prior models for smooth rotation and expansion using techniques similar to those in the slow-and-smooth model [23] (e.g. Green functions of differential operators). The theory gives good agreement with the trends observed in four human experiments.

1 Introduction

As an observer moves through the environment, the retinal image changes over time to create multiple complex motion flows, including translational, circular and radial motion. Imagine that you are in a moving car and seeing a walker through a set of punch holes, as illustrated in figure (1). In each hole, the component of the line motion orthogonal to the lines orientation can be clearly perceived; however, the component of motion parallel to the lines orientation is not observable. This is often referred to as the aperture problem. The inherent ambiguity of motion signals requires the visual system to employ an efficient integration strategy to combine many local measurements in order to perceive global motion. In the example (figure (1)), the visual system needs to effectively integrate local signals to perceive multiple complex motion flows, including the movement of the walker and the moving background respectively.

Human observers are able to process different motion patterns and infer ego motion and global structure of the world. Psychophysical experiments have identified a variety of phenomena, such as motion capture and motion cooperativity [16], which appear to be consequences of such integration. A number of computational Bayesian models have been proposed to explain these effects based on prior assumptions about motion. In particular, it has been shown that a slow-and-smooth prior, and related models, can qualitatively account for a range of experimental results [23, 21, 22] and can quantitatively account for others [9, 17].

However, the integration strategy modeled by the slow-and-smooth prior may not generalize to more complex motion types, such as circular and radial motion, which are critically important for estimating ego motion. In this paper we

are concerned with two questions. (1) What integration priors should be used for a particular motion input? (2) How can local motion measurements be combined with the proper priors to estimate motion flow? Within the framework of Bayesian inference, the answers to these two questions are respectively based on model selection and parameter estimation. In the field of motion perception, most work has focused on the second question, using parameter estimation to estimate motion flow. However, Stocker and Simoncelli [18] recently proposed a conditioned Bayesian model in which strong biases in precise motion direction estimates arise as a consequence of a preceding decision about a particular hypothesis (left vs. right motion).
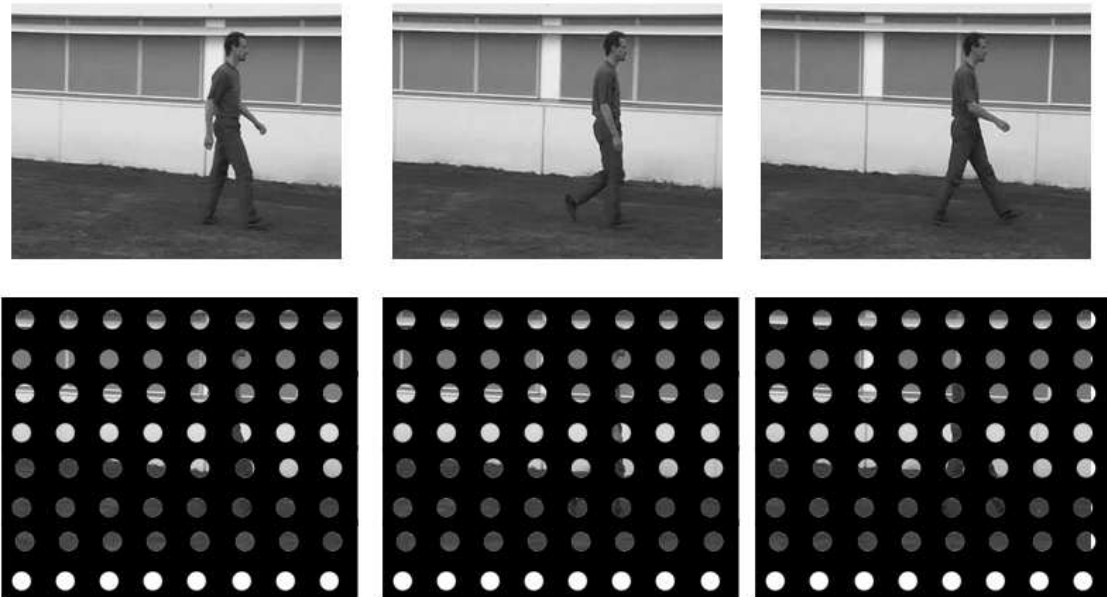
Fig. 1. An illustration of observing a walker with a moving camera. Top panel, three example frames. Bottom panel, observing the scene through a set of punch holes

The goal of this paper is to provide a computational explanation for both of the above questions using Bayesian inference. To address the first question, we develop new prior models for smooth rotation and expansion motion. To address the second, we propose that the human visual system has available multiple models of motion integration appropriate for different motion patterns. The visual system decides the best integration strategy based upon the perceived motion information, and this choice in turn affects the estimation of motion flow.

In this paper, we first present a computational theory in section (4) that includes three different integration strategies, all derived within the same framework. We test this theory in sections (5,6) by comparing its predictions with human performance in psychophysical experiments, in which subjects were asked

to detect global motion patterns or to discriminate motion direction in translation (left vs. right), rotation (clockwise vs. counter-clockwise), and radial motion (inward vs. outward). We employ three commonly used stimuli: random dot patterns, moving gratings and plaids (two superimposed gratings each with different orientations) to show that the model can be applied to a variety of inputs. We compare the predictions of the model predictions with human performance across four psychophysical experiment.

2   Background

There is an enormous literature on visual motion phenomena and there is only room to summarize the work most relevant to this paper. Our computational model relates most closely to work [23, 21, 9] that formulates motion perception as Bayesian inference with a prior probability biasing towards slow-and-smooth motion. But psychophysical [4, 10, 1, 8], physiological [19, 3] and fMRI data [11] suggests that humans are sensitive to a variety of motion patterns including translation, rotation, and expansion. In particular, Freeman *et al* [4] and Lee *et al* [8] demonstrated that human performance on discrimination tasks for translation, rotation, and expansion motion was inconsistent with the predictions of the slow-and-smooth theory (our simulations independently verify this result). Instead, we propose that human motion perception is performed at two levels of inference: (i) model selection, and (ii) estimating the velocity with the selected model. The concept of model selection has been described in the literature, see [7], but has only recently been applied to model motion phenomena [18]. Our new motion models for rotation and expansion are formulated very similarly to the original slow-and-smooth model [23] and similar mathematical analysis [2] is used to obtain the forms of the solutions in terms of Greens functions of the differential operators used in the priors.

Smoothness priors for motion have a long history in computer vision beginning with the influential work of Horn and Schunk [6] and related work by Hildreth [5]. The slow-and-smooth prior on the velocity [23] was developed to explain psychophysical phenomena such as motion capture [13]. The nature of the psychophysics stimuli and the experimental phenomena meant that the first derivative regularizers used in [6][5] needed to be supplemented by zeroth order regularizers (i.e. slowness of the velocity) and second, and higher, order regularizers. Specifically, higher order regularizers were required to ensure that the theory was well-posed for the sparse stimuli (e.g. moving dots) used in typical psychophysical experiments. Moreover, zeroth order regularizers were needed to ensure that the interactions (e.g. motion capture) between different stimuli decreased with distance [24]. Justifications for the slowness and smoothness assumptions came from analyzes of the probability distribution of velocity flow in the image plane assuming isotropic distributions of velocity in space [20, 25] which showed that for most distributions of velocity in space the projection onto the image plane would induce velocity distributions which were peaked at zero velocity and zero velocity gradient. More recently, empirical studies by Roth and

Black [15] on motion statistics yield similar results. We note that a limitation of the slow-and-smooth prior, for computer vision purposes, is that it is a quadratic regularizer and hence tends to smooth the velocity over motion discontinuities (although the limited range of interaction of the slow-and-smooth model means that this effect is more localized than for standard quadratic regularizers). Hence non-quadratic regularizers/priors are preferred for computer vision applications. The nature of the psychophysical stimuli in this paper, however, means that discontinuities can be ignored and so we can use quadratic regularizers and take advantage of their good computational properties.

In order to relate theories like slow-and-smooth to psychophysics we have to understand the experimental design and model how human subjects address the specific experimental tasks. The experiments we describe in this paper use a standard experimental design which we describe for dot stimuli (a similar design is used for gratings and plaid stimuli). For these stimuli, some of the dots (the signal) move coherently, whereas other dots (the noise) move randomly. The experimental task is to measure some property of the coherent stimulus (e.g. does it move to the left or the right? does it rotate or contract?). The difficulty of performing this task depends on the proportion of signal dots to noise dots. The *coherence threshold* is defined to be the value of the ratio at which human subjects achieve 75 % accuracy for the task. Hence the psychophysicist can adjust the parameters of the stimuli (i.e. proportion of signal to noise) until this degree of accuracy is obtained. To evaluate the performance of a theory, we compute the analogous threshold for the theory and compare it to the threshold found by the experiments. An alternative criterion is to measure the minimum speed or contrast (e.g. brightness of the dots compared to the background) required to achieve a certain performance level (e.g. 75 % accuracy) at a visual task (e.g. is the motion to the right or the left?).

3   Overview of the Theory

We formulate motion perception as a problem of Bayesian inference with two stages followed by post-processing. We hypothesize that the human visual system has a set of models M for estimating the velocity. Firstly, we perform model selection to find the model that best explains the observed motion pattern. Secondly, we estimate motion properties using the selected model. The post-processing models how human subjects use the estimated motion to perform psychophysical tasks.

Formally a model M estimates the velocity field $\{v\}$ defined at all positions $\{r\}$ in the image from local velocity measurements $\{u\}$ at discrete positions $\{r_i,\ i = 1, \ldots N\}$. The model is specified by a prior $p(\{v\}|M)$ on the velocity field and a likelihood function $p(\{u\}|\{v\})$ for the measurements. This specifies a full distribution $p(\{v\}, \{u\}, M) = p(\{u\}|\{v\})p(\{v\}|M)P(M)$, where $P(M)$ is the prior over the models (which is a uniform distribution).

The prior and likelihoods are expressed as Gibbs distributions:

$$p(\{\boldsymbol{v}\}|M) = \exp(-E(\{\boldsymbol{v}\}|M)/T), \tag{1}$$

$$p(\{\boldsymbol{u}\}|\{\boldsymbol{v}\}) = \exp(-E(\{\boldsymbol{u}\}|\{\boldsymbol{v}\})/T), \tag{2}$$

and are specified in sections (4.1,4.2) respectively.

The first stage of the theory selects the model M^* that best explains the measurements $\{\boldsymbol{u}\}$. It is chosen to maximize the posterior of the model evidence:

$$M^* = \arg\max_M P(M|\{\boldsymbol{u}\}) = \arg\max_M \frac{P(\{\boldsymbol{u}\}|M)P(M)}{P(\{\boldsymbol{u}\})} \tag{3}$$

$$\text{where} \quad p(\{\boldsymbol{u}\}|M) = \int p(\{\boldsymbol{u}\}|\{\boldsymbol{v}\})p(\{\boldsymbol{v}\}|M)d\{\boldsymbol{v}\}. \tag{4}$$

The second stage of the theory estimates the velocity flow by using the selected model M^*. This yields:

$$\boldsymbol{v}^* = \arg\max_{\boldsymbol{v}} p(\{\boldsymbol{v}\}|\{\boldsymbol{u}\}, M^*) \tag{5}$$

$$\text{where} \quad p(\{\boldsymbol{v}\}|\{\boldsymbol{u}\}, M^*) = \frac{p(\{\boldsymbol{u}\}|\{\boldsymbol{v}\})p(\{\boldsymbol{v}\}|M^*)}{p(\{\boldsymbol{u}\}|M^*)}. \tag{6}$$

The form of these priors $P(\{\boldsymbol{v}\})$ are distributions over functions, see section (4.1), which means that this model evidence will be computed by techniques similar to those used for Gaussian Processes [14].

The post-processing estimates the properties of the motion flow required to perform the psychophysics tasks. In the psychophysics experiments in this paper, the human subjects must determine whether the motion is translation, expansion/contraction, or rotation and then estimate the appropriate motion properties: (i) direction of motion (if translation), (ii) direction of expansion/contraction (if expansion/contraction), or (iii) direction of rotation (if rotation).

To address these tasks, we use the following post-processing strategy which we illustrate for the rotation stimuli. First we compute the two stages of the theory as described above. If the rotation model is selected, then we estimate the center of rotation (by least squares fit) and compute the direction of rotation (i.e. clockwise or counterclockwise) for the velocity field evaluated at each data point (i.e. dots), see Appendix (C) for details. We alter the parameters of the stimuli (i.e. the proportion of noise dots) until 75 % of the stimulus data points are estimated to have the correct direction of rotation (for each stimulus). Alternatively, we consider a set of stimuli, compute the velocity flow for each stimulus and estimate the centers of rotation, compute the histograms of the estimated rotated directions for the data points for all images, and determine the stimulus parameters so that 75 % of the histogram values are in the correct direction.

4   Model Formulation

This section specifies the priors for the different models, the likelihood functions, closed form solutions for the velocity flows, and analytic formulae for performing model selection. Some details of the derivations are described in Appendix (A).

4.1   The Priors

We define three priors corresponding to the three different types of motion –
translation, rotation, and expansion. For each motion type, we encourage slow-
ness and spatial smoothness. The prior for translation is very similar to the
slow-and-smooth prior [23] except we drop the higher-order derivative terms
and introduce an extra parameter (to ensure that all three models have similar
degrees of freedom).

   We define the priors by their energy functions $E(\{\boldsymbol{v}\}|M)$, see equation (1).
We label the models by $M \in \{t, r, e\}$, where t, r, e denote translation, rotation,
and expansion respectively. (We note that the prior for expansion will also ac-
count for contraction). We use the convention that $\boldsymbol{v} = (v_x, v_y)$.

1. Slow-and-smooth-translation:

$$E(\{\boldsymbol{v}\}|M = t) = \int \lambda(|\boldsymbol{v}|^2 + \mu|\nabla\boldsymbol{v}|^2 + \eta|\nabla^2\boldsymbol{v}|^2)d\boldsymbol{r}, \tag{7}$$

2. Slow-and-smooth-rotation:

$$E(\{\boldsymbol{v}\}|M = r) = \int \lambda\{|\boldsymbol{v}|^2 + \mu[(\frac{\partial v_x}{\partial x})^2 + (\frac{\partial v_y}{\partial y})^2 + (\frac{\partial v_x}{\partial y} + \frac{\partial v_y}{\partial x})^2] + \eta|\nabla^2\boldsymbol{v}|^2\}d\boldsymbol{r}, \tag{8}$$

3. Slow-and-smooth-expansion:

$$E(\{\boldsymbol{v}\}|M = e) = \int \lambda\{|\boldsymbol{v}|^2 + \mu[(\frac{\partial v_x}{\partial y})^2 + (\frac{\partial v_y}{\partial x})^2 + (\frac{\partial v_x}{\partial x} - \frac{\partial v_y}{\partial y})^2] + \eta|\nabla^2\boldsymbol{v}|^2\}d\boldsymbol{r}, \tag{9}$$

   These priors are motivated as follows. The $|\boldsymbol{v}|^2$ and $|\nabla^2\boldsymbol{v}|^2$ terms bias towards
slowness and smoothness and are used in all three models. However, the priors
differ in the use of first derivative terms. The translation prior prefers constant
translation motion with $\boldsymbol{v}$ constant, since $\nabla\boldsymbol{v} = 0$ for this type of motion. The
translation prior is similar to the first three terms of the slow-and-smooth energy
function [23] but with a restriction on the set of parameters. Formally $\lambda(|\boldsymbol{v}|^2 +
\frac{\sigma^2}{2}|\nabla\boldsymbol{v}|^2 + \frac{\sigma^4}{8}|\nabla^2\boldsymbol{v}|^2)d\boldsymbol{r} \approx \lambda\sum_{m=0}^{\infty}\frac{\sigma^{2m}}{m!2^m}|D^m\boldsymbol{v}|^2d\boldsymbol{r}$ (for appropriate parameter
values). Our computer simulations showed that the translation prior performs
similar to the original slow-and-smooth prior.

   The first derivative terms in the slow-and-smooth-rotation prior is motivated
by the ideal form of rigid rotation:

$$v_x = -\omega(y - y_0), \quad v_y = \omega(x - x_0) \tag{10}$$

where (x_0, y_0) are the (unknown) centers, ω is the (unknown) angular speed.
This rigid rotation is preferred by the slow-and-smooth-rotation prior (at least,
by its first derivative terms), since we have $\frac{\partial v_x}{\partial y} + \frac{\partial v_y}{\partial x} = 0$ and $\frac{\partial v_x}{\partial x} = \frac{\partial v_y}{\partial y} = 0$
(independent of (x_0, y_0) and ω).

The first derivative terms in the slow-and-smooth-expansion prior is motivated by the ideal form of rigid expansion:

$$v_x = e(x - x_0), \quad v_y = e(y - y_0) \tag{11}$$

where (x_0, y_0) are the (unknown) centers, and e is the (unknown) expansion rate. Such a rigid expansion is preferred by the slow-and-smooth-expansion prior (first order terms), since we have $\frac{\partial v_x}{\partial x} - \frac{\partial v_y}{\partial y} = 0$, and $\frac{\partial v_x}{\partial y} = \frac{\partial v_y}{\partial x} = 0$ (again independent of (x_0, y_0) and e).

Note that we will use equations (10,11) in the post-processing stage where we estimate the centers (x_0, y_0) and the rotation/expansion ω, e, see Appendix (C).

4.2   The Likelihood Functions

The likelihood function differs for the three classes of stimuli used in the psychophysics experiments that we considered: (i) For moving dots, as used in [4], the motion measurement is the local velocity $\boldsymbol{u}$ of each dot (assumed known since the motion is small and so correspondence between dots at different time frames is unambiguous); (ii) For a moving grating [12] [8], the local motion measurement is the velocity component in the direction orthogonal to the grating orientation; (iii) For a moving plaid (with two superimposed grating each with different orientation) [12] [8], the local motion measurement uses the velocity components of each grating (effectively this reduces to the moving dot case).

For the dot stimuli, the energy term in the likelihood function, see equation (2), is set to be

$$E(\{\boldsymbol{u}|\boldsymbol{v}\}) = \sum_{i=1}^{N} |\boldsymbol{v}(\boldsymbol{r}_i) - \boldsymbol{u}(\boldsymbol{r}_i)|^2 \tag{12}$$

For the grating stimuli, see figure (5), the likelihood function uses the energy function

$$E_n(\{\boldsymbol{u}\}|\{\boldsymbol{v}\}) = \sum_{i=1}^{N} |\boldsymbol{v}(\boldsymbol{r}_i) \cdot \hat{\boldsymbol{u}}(\boldsymbol{r}_i) - |\boldsymbol{u}(\boldsymbol{r}_i)||^2 \tag{13}$$

where $\hat{\boldsymbol{u}}(\boldsymbol{r}_i)$ is the unit vector in the direction of $\boldsymbol{u}(\boldsymbol{r}_i)$ (i.e. in the direction of local image gradient and hence perpendicular to the orientation of the bar).

For the plaid stimuli, see figure (6), the likelihood function uses the energy function

$$E_p(\{\boldsymbol{u}\}|\{\boldsymbol{v}\}) = \sum_{i=1}^{N} (|\boldsymbol{v}(\boldsymbol{r}_i) \cdot \hat{\boldsymbol{u}}_1(\boldsymbol{r}_i) - |\boldsymbol{u}_1(\boldsymbol{r}_i)||^2 + |\boldsymbol{v}(\boldsymbol{r}_i) \cdot \hat{\boldsymbol{u}}_2(\boldsymbol{r}_i) - |\boldsymbol{u}_2(\boldsymbol{r}_i)||^2) \tag{14}$$

where there are two sets of measurements $\{u_1\}$ and $\{u_2\}$. This specifies the velocity in two different directions and hence can be reformulated in terms of model for the dots.

4.3   MAP estimator of velocities

The MAP estimate of the velocities for each model is obtained by solving

$$\boldsymbol{v}^* = \arg\max_{\boldsymbol{v}} p(\{\boldsymbol{v}\}|\{\boldsymbol{u}\}, M) = \arg\min_{\boldsymbol{v}}\{E(\{\boldsymbol{u}\}|\{\boldsymbol{v}\}) + E(\{\boldsymbol{v}\}|M)\} \qquad (15)$$

For the slow-and-smooth model [23], it was shown using regularization analysis [2] that this solution $\boldsymbol{v}^*$ can be expressed in terms of a linear combination of the Green function G of the differential operator which imposes the slow-and-smoothness constraint (the precise form of this constraint was chosen so that G was a Gaussian [24]).

In this paper, we perform a similar analysis for the three types of models $M \in \{t, r, e\}$ introduced in this paper and show that the solution $\boldsymbol{v}^*$ can also be expressed as a linear combination of Green functions G^M where M indicates the model. The precise forms of these Green functions are determined by the differential operators associated with the slow-and-smoothness terms of the models and are derived in Appendix (A). The main difference with the Green functions described by Yuille and Grzywacz [23],[25] is that the models for expansion/contraction and rotation require two vector valued Green functions $\boldsymbol{G}_1^M = (G_{1x}^M, G_{1y}^M)$ and $\boldsymbol{G}_2^M = (G_{2x}^M, G_{2y}^M)$ (with the constraint that $G_{1x}^M = G_{2y}^M$ and $G_{2x}^M = G_{1y}^M$). These vector-valued Green functions are required to perform the coupling between the different velocity component required for rotation and expansion, see figure (2). For the translation model there is no coupling required and so $G_{2x}^M = G_{1y}^M = 0$.
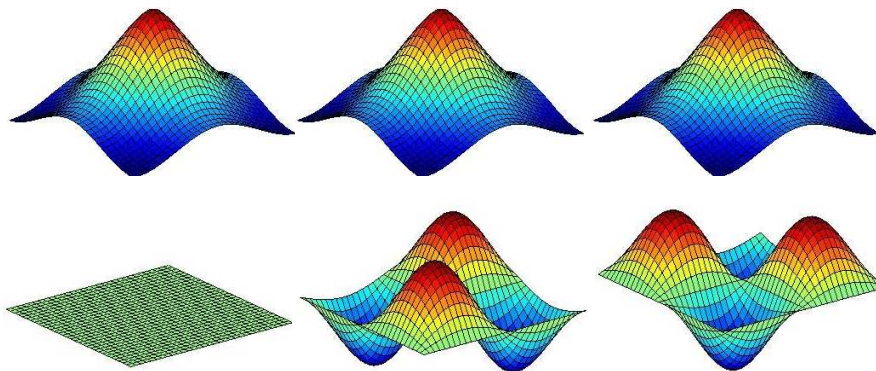
Fig. 2. The vector-valued Green function $\boldsymbol{G} = (G_1, G_2)$. Top panel, left-to-right: $G_{1x}^{M=t}, G_{1x}^{M=r}, G_{1x}^{M=e}$ for the translation, rotation, and expansion models. Bottom panel: left-to right: $G_{2x}^{M=t}, G_{2x}^{M=r}, G_{2x}^{M=e}$ for translation, rotation, and expansion models. Observe that the G_{1x}^M are similar for all models, $G_{2x}^{M=t}$ vanishes for the translation model (i.e. no coupling between velocity components), and $G_{2x}^{M=r}$ and $G_{2x}^{M=e}$ both have two peaks which correspond to the two directions of rotation and expansion. Recall that $G_{1x}^M = G_{2y}^M$ and $G_{2x}^M = G_{1y}^M$.

Our analysis, see Appendix (A), shows that the estimated velocities for model M can be expressed in terms of the Greens functions $\boldsymbol{G}_1^M, \boldsymbol{G}_2^M$ of the model and coefficients $\{\alpha_i\}, \{\beta_i\}$:

$$\boldsymbol{v}(\boldsymbol{r}) = \sum_{i=1}^N [\alpha_i \boldsymbol{G}_1^M (\boldsymbol{r} - \boldsymbol{r}_i) + \beta_i \boldsymbol{G}_2^M (\boldsymbol{r} - \boldsymbol{r}_i)], \tag{16}$$

where the coefficients $\{\alpha_i\}, \{\beta_i\}$ are obtained by solving linear equations, derived in Appendix (A), which depend on the likelihood function and hence on the form of the stimuli (e.g. dot, grating, or plaid).

For *the dot stimuli*, the $\{\alpha\}, \{\beta\}$ are obtained by solving the linear equations:

$$\sum_{i=j}^N [\alpha_j \boldsymbol{G}_1^M (\boldsymbol{r}_i - \boldsymbol{r}_j) + \beta_j \boldsymbol{G}_2^M (\boldsymbol{r}_i - \boldsymbol{r}_j)] + \alpha_i \boldsymbol{e}_1 + \beta_i \boldsymbol{e}_2 = \boldsymbol{u}(r_i), \ i = 1, \ldots N, \tag{17}$$

where $\boldsymbol{e}_1, \boldsymbol{e}_2$ denote the (orthogonal) coordinate axes (i.e. the x and y axes). If we express the $\{\alpha\}, \{\beta\}$ as two N-dimensional vectors A and B (recall N is the number of dots), the $\{u_x\}$ and $\{u_y\}$ as vectors $U = (U_x, U_y)$, and define $N \times N$ matrices $g_{1x}^M, g_{2x}^M, g_{1y}^M, g_{2y}^M$ to have components $G_{1x}^M(\boldsymbol{r}_i - \boldsymbol{r}_j), G_{2x}^M(\boldsymbol{r}_i - \boldsymbol{r}_j), G_{1y}^M(\boldsymbol{r}_i - \boldsymbol{r}_j), G_{2y}^M(\boldsymbol{r}_i - \boldsymbol{r}_j)$ respectively, then we can express these linear equations as:

$$[\begin{pmatrix} g_{1x}^M & g_{2x}^M \\ g_{1y}^M & g_{2y}^M \end{pmatrix} + I] \begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} U_x \\ U_y \end{pmatrix} \tag{18}$$

For *the grating stimuli*, the $\{\alpha\}, \{\beta\}$ satisfy similar linear equations:

$$[\begin{pmatrix} \tilde{g}_{1x}^M & \tilde{g}_{2x}^M \\ \tilde{g}_{1y}^M & \tilde{g}_{2y}^M \end{pmatrix} + I] \begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} U_x \\ U_y \end{pmatrix} \tag{19}$$

where $\tilde{g}_{1x}^M$ is the matrix with components $\tilde{G}_{1x}^M(\boldsymbol{r}_i - \boldsymbol{r}_j) = [\boldsymbol{G}_1^M(\boldsymbol{r}_i - \boldsymbol{r}_j) \cdot \hat{\boldsymbol{u}}(r_i)] \hat{\boldsymbol{u}}_x(r_i)$, and similarly for $\tilde{g}_{1y}^M, \tilde{g}_{2x}^M$ and $\tilde{g}_{2y}^M$.

For *the plaid stimuli*, the $\{\alpha\}, \{\beta\}$ satisfy:

$$(\begin{pmatrix} \tilde{g}_{1x}^M & \tilde{g}_{2x}^M \\ \tilde{g}_{1y}^M & \tilde{g}_{2y}^M \end{pmatrix}_{u_1} + \begin{pmatrix} \tilde{g}_{1x}^M & \tilde{g}_{2x}^M \\ \tilde{g}_{1y}^M & \tilde{g}_{2y}^M \end{pmatrix}_{u_2} + I) \begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} U_{1x} \\ U_{1y} \end{pmatrix} + \begin{pmatrix} U_{2x} \\ U_{2y} \end{pmatrix}, \tag{20}$$

where U_1, U_2 are the velocity components of the two gratings respectively.

These equations (18,19,20) can all be solved by matrix inversion to determine the coefficients $\{\alpha\}, \{\beta\}$. These solutions can be substituted into equation (16) to obtain the velocity field.

4.4   Model Selection

We re-express model evidence $p(\{\boldsymbol{u}\}|M)$ in terms of (A, B):

$$p(\{\boldsymbol{u}\}|M) = \int p(\{\boldsymbol{u}\}|A, B, M) p(A, B|M) dA dB, \tag{21}$$

where $p(\{\boldsymbol{u}\}|A, B, M), p(A, B|M)$ are obtained by substituting the form of the velocity estimate, equation (16), and then normalizing these distributions. This is similar to how analogous terms are calculated for Gaussian Processors [14]. It can also be considered a variant of the Laplace approximation. The analysis is given in Appendix (B) and shows that the model evidence can be computed analytically be exploiting properties of multi-dimensional Gaussians. The result can be expressed compactly by introducing new notation in the form of $2N \times 2N$ matrices:

$$
g^M = \begin{pmatrix} g_{1x}^M & g_{2x}^M \\ g_{1y}^M & g_{2y}^M \end{pmatrix} \quad \tilde{g}^M = \begin{pmatrix} \tilde{g}_{1x}^M & \tilde{g}_{2x}^M \\ \tilde{g}_{1y}^M & \tilde{g}_{2y}^M \end{pmatrix}.
$$

For *the dot stimuli* this gives:

$$
p(\{\boldsymbol{u}\}|M) = \frac{1}{(\pi T)^N \sqrt{\det(g^M + I)}} \exp[-\frac{1}{T}(U^T U - U^T \frac{g^M}{g^M + I} U)] \qquad (22)
$$

Similarly, for *the gratings stimuli* we obtain:

$$
p(\{\boldsymbol{u}\}|M) = \frac{1}{(\pi T)^N} \frac{\sqrt{\det(g^M)}}{\sqrt{\det(\tilde{\Sigma})}} \exp[-\frac{1}{T}(U^T U - U^T \tilde{g}^M \tilde{\Sigma}^{-1} \tilde{g}^M U)] \qquad (23)
$$

where $\tilde{\Sigma} = \tilde{g}^M \tilde{g}^M + g^M$.

For the *plaids stimuli* we obtain: (we omit the M in g^M to simplify notation.)

$$
p(\{\boldsymbol{u}_1\}, \{\boldsymbol{u}_2\}|M) = \frac{1}{(\pi T)^N} \frac{\sqrt{\det(g)}}{\sqrt{\det(\tilde{\Sigma})}} \exp -\frac{1}{T}[U_1^T U_1 + U_2^T U_2 - \Phi^T \tilde{\Sigma}^{-1} \Phi] \quad (24)
$$

where $\tilde{\Sigma} = \tilde{g}_1^M \tilde{g}_1^M + \tilde{g}_2^M \tilde{g}_2^M + g^M$ and $\Phi = \tilde{g}_1^T U_1 + \tilde{g}_2^T U_2$.
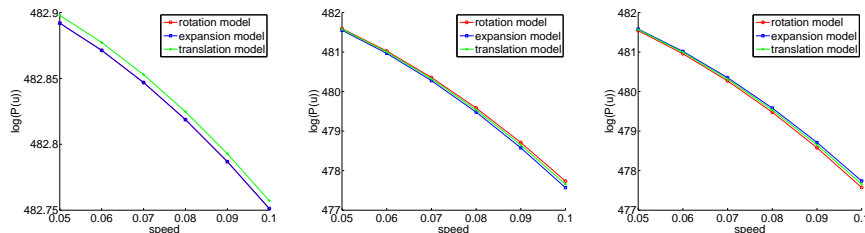
5   Experiment 1: random dot motion

We first investigate motion perception with the moving dots stimuli (figure 3) used by Freeman and Harris [4]. The stimuli consisted of 128 moving dots in a display window. All the dots in the stimulus had the same speed in all the three motion patterns, including translation, rotation and expansion.

The reported simulations are based upon 1000 trials for each speed condition. The parameter values used in the simulations are $\lambda = 0.001$, $\mu = 12.5$, $\eta = 78.125$ and $T = 0.0054$.

Model selection results are shown in the panels of figure (4). It is shown that the correct model is always selected over the entire range of speed, and for all 3 type of motion stimuli.
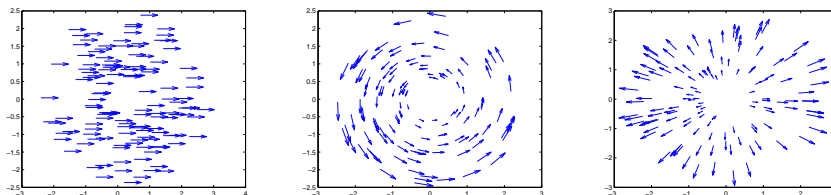
Fig. 3. Moving random dot stimuli – translation, rotation and expansion

Fig. 4. Model selection results in Experiment 1 with random dot motion. Plots the log probability of the model as a function of speed for each type of stimuli. left: translation; middle: rotation; right: expansion. Green curves with cross are from translation model. Red curves with circles are from rotation model. Blue curves with squares are from expansion model.

6   Experiment 2: randomly oriented gratings

6.1   Stimulus

When randomly oriented grating elements drift behind apertures, the perceived direction of motion is heavily biased by the orientation of the gratings, as well as by the shape and contrast of the apertures. Recently, Nishida and his colleagues developed a novel global motion stimulus consisting of a number of gratings elements, each with randomly assigned orientation [12]. Lee and his colleagues [8] extended this stimulus to rotational and radial motion. A coherent motion in this type of stimuli is perceived when the drifting velocities, also termed the component of velocity, of elements are consistent with a given global motion pattern. Examples of the stimuli used in these psychophysical experiments are shown in left side of figure (5). The stimuli consisted of 728 gratings (drifting sine-wave gratings windowed by stationary Gaussians). The orientations of the gratings were randomly assigned, and their drifting velocities were determined by a specified global motion. The motions of signal grating elements were consistent with global motion, but the motions of noise grating elements were randomized. The task was to identify the global motion direction as one of two alternatives: left/right for translation, clockwise/counterclockwise for rotation, and inward/outward for expansion. Motion sensitivity was measured by the coherence threshold, defined as the proportion of signal elements that yielded a performance level of 75% correct.

Similar stimuli with 328 gratings were generated to test our computational models. The smaller grating number is used to shorten simulation time, and experiments show that model behavior remains the same when grating number increases. The input for the models is the component velocity perpendicular to the assigned orientation for each grating. Two levels of inference are involved. Our simulations first select the correct model for each stimulus and then estimate the velocity flow using the selected model.

All reported simulations in Experiment 2-4 are based up 50 trials for each level of coherence ratio. The parameter values used in the simulations are $\mu = 12.5$, $\eta = 78.125$, $v = 0.0244$ and $T = 4 \exp -5.0$. For rotation/expansion model, $\lambda = 0.005$; for translation model, $\lambda = 0.0025$.

6.2   Model selection results

Model selection results are shown in tables (1,2,3). From the table we see that the rotation model has the highest model evidence for the rotation stimulus over the entire range of the coherence ratio. As coherence ratio increases, the advantage of rotation model over the other two models also increases. Analogous results are found for the expansion stimuli. However, for translation stimulus, all models have very similar model evidence values. This is caused by translation being favored by all three models, and indeed they all produce very similar motion field estimates.

6.3   Comparison between human performance and model prediction

Once the appropriate model has been selected, the velocity flow is estimated using the selected model (more specifically, this is computed from equations (16,17) with the appropriate M).

The post-processing in Experiments 2 4 includes a simple decision rule the input of which is the velocity flow estimated by the selected model.First the selected model estimates velocity flow in each trial. At each level of coherence ratio in the range of 0 to 50%, the post-processing enables the model to predict accuracy in discriminating motion direction for each stimulus (e.g., left vs. right for translation, clockwise vs. counter-clockwise for rotation, and inward vs. outward for expansion). To address this task, we pool estimated motion directions of 328 elements to generate a motion direction histogram for each trial. Note we use rotation direction and expansion direction to generate histograms. We then compute the average direction histogram over 50 trials for stimuli with one moving direction (e.g., $P(\boldsymbol{v}|L)$ for leftward translation), and the average direction histogram over 50 trials for stimuli with the opposite moving direction, $P(\boldsymbol{v}|R)$ for rightward translation. Based on the two computed histograms, a decision boundary is searched to achieve maximum accuracy. We apply the same post-processing on each of the three motion patterns to estimate accuracy at each level of coherence ratio, and then estimate the coherence threshold needed to achieve 75% correct in the discrimination task.

coherence ratio	10%	20%	30%	40%	50%
rotation model	1009.0	1060.1	1146.3	1256.9	1418.8
expansion model	1008.9	1059.5	1144.9	1254.4	1414.7
translation model	1008.9	1059.7	1145.5	1255.5	1416.5

Table 1. Model selection result for the grating rotation stimuli. The values are logarithms of model evidence. The correct model, rotation model, always wins. As the coherence ratio increases, the rotation model's advantage also increases.

coherence ratio	10%	20%	30%	40%	50%
rotation model	1011.5	1057.4	1151.8	1275.9	1430.0
expansion model	1011.7	1058.0	1153.4	1278.5	1434.2
translation model	1011.6	1057.6	1152.5	1277.1	1431.9

Table 2. Model selection result for the grating expansion stimuli. The correct model, expansion, always wins.

coherence ratio	10%	20%	30%	40%	50%
rotation model	990.2	1076.3	1159.7	1265.6	1456.1
expansion model	990.2	1076.2	1159.7	1265.7	1456.3
translation model	990.2	1076.2	1159.7	1265.6	1456.2

Table 3. Model selection result for the grating translation stimuli. All three models (rotation/expansion/translation) have virtually the same model evidence. This is due to the fact that rotation/expansion models also favor translation as translation model does.

The results of psychophysical experiments (middle panel of figure 5) showed worse performance for perceiving translation than rotation/expansion motion [8]. Clearly, as shown in the third panel of the same figure, the model performs best for rotation and expansion, and is worst for translation. This finding agrees with human performance in psychophysical experiments.
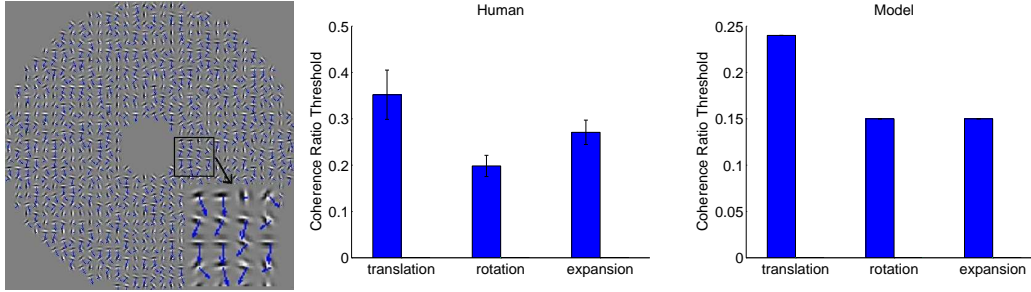
Fig. 5. Stimulus and results in Experiment 2 with randomly-oriented grating stimuli. Left panel: illustration of grating stimulus. Blue arrows indicate the drifting velocity of each grating. Middle panel: human coherence thresholds for different motion stimuli. Right panel: Model prediction of coherence thresholds which are consistent with human trends.
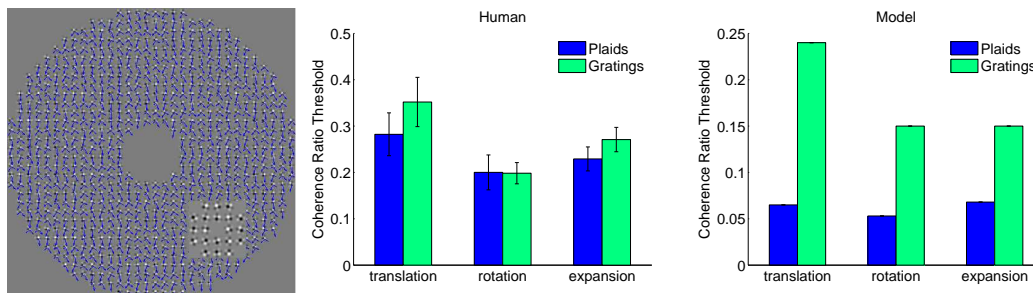
7   Experiment 3: plaid stimuli

7.1   Stimulus

A plaid pattern consists of two drifting gratings arranged such that their individual orientations are orthogonal (see illustration in the left panel of figure 6). The drifting velocities of the two gratings in one plaid are determined by the global motion and their orientations. As the point of intersection for two superimposed gratings provides a tracking feature, the plaid stimulus reduces ambiguity in the measurement of local motion. The experimental methods were the same as those used in Experiment 2, except plaid elements were replaced by grating elements.

7.2   Results

As shown in the middle panel of figure 6, human observers exhibited the worst performance (highest thresholds) in discriminating motion directions in translation, as compared to rotation and radial motion. The right panel of figure 6 shows the model's performance, which matches human performance quantitatively. Notably, the model predicts better performance for plaid stimuli (Exp 3) than for grating stimuli (Exp 2), which is consistent with human performance.

Fig. 6. Stimulus and results in plaid experiment in Experiment 3 with plaid stimuli. Left panel: illustration of plaid stimulus. Blue arrows indicate the drifting velocity of each grating. Middle panel: human coherence thresholds for different motion stimuli. Right panel: Model prediction of coherence thresholds which are consistent with human trends.

8 Experiment 4: Results for non-rigid rotation and radial motion

8.1 Stimulus

In each of the previous three experiments, all the elements were assigned the same speed, which yielded non-rigid rotation and radial motion. However, in a rigid motion pattern, the assigned speed for each element should vary according to the distance to the center of rotation and radial motion. In Experiment 4, we assessed whether rigidity could affect human motion performance in perceiving rotation and radial motion. We introduced a rigid rotation condition, in which element speeds were determined by a constant rotation velocity and the distance of the element to the rotation center. The average speed was kept the same as that used in non-rigid rotation condition. Similar manipulations were employed for the rigid radial-motion condition. The experimental methods were the same as those employed in Experiment 2 with the grating stimulus.
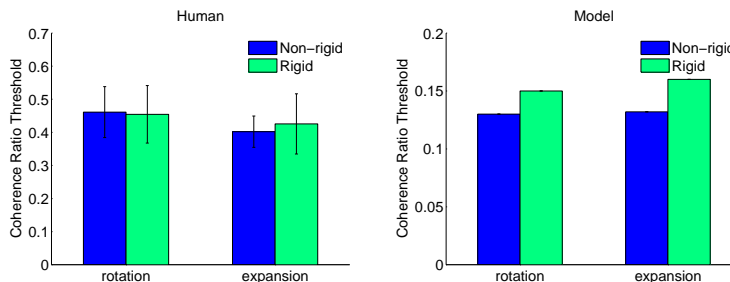
8.2 Results

As shown in figure 7 (left panel), human observers were not sensitive to rigid versus non-rigid rotation and radial motion (in the current experimental conditions). This result is consistent with previous findings reported in the literature [1]. The right panel in figure 7 shows the model's predictions, which qualitatively match human performance across the four conditions. Although the rotation and expansion priors were motivated by rigid rotation and expansion, the results in Experiment 4 clearly the priors are "robust" to non-rigid

9 Conclusion

We propose a computational theory for motion perception which proceeds in two stages. The first stage selects the best model to explain the data from a

Fig. 7. Results in Experiment 4 with grating stimuli to compare rigid versus nonrigid rotation and expansion. Left panel: human coherence thresholds for rigid and non-rigid conditions as a function of different motion patterns. Right panel: Model prediction of coherence thresholds which are consistent with human trends.

set of competing models. The second stage estimates the velocity field using the selected model. In this paper, we used three competing models for translation, rotation, and expansion. We defined novel priors for rotation and expansion by analogy to the slow-and-smooth prior proposed by Yuille and Grzywacz [23, 24] and showed that the velocity estimation could be expressed as a linear combination of Green functions. We showed that model selection correctly selects the appropriate model for several different stimulus conditions and that the selected model estimates a plausible velocity field.

We compared our *competitive prior* theory to the performance of human subjects on threshold tasks using different types of stimuli. Our results on random dot stimuli were in general agreement with the experimental findings of Freeman and Harris [4]. We also found good agreement with the experimental findings of our group [8] for grating and plaid stimuli. But there are current limitations to our results since different parameter settings are required for the model estimation and the velocity estimation tasks. We are exploring two possibilities: (i) for the psychophysics stimuli the human subjects are aware that the centers of expansion and rotation are at the center of the image, but this information is not currently being used in our theory (but can be inserted as an additional requirement), (ii) humans may be better at observing expansion/rotation than translation since attention and eye movements (A. Glennerster – private communication) may be used to remove the translational components leaving only expansion/rotation, but this is not included in our model.

Our current work aims to resolve these issues and extend these findings to a range of different motions (e.g. affine motion) and to use increasingly naturalistic appearance/intensity models. It is also important to determine if motion patterns to which humans are sensitive correspond to those appearing regularly in natural motion sequences.

Appendix A

This appendix gives the derivations of the formulae for estimating velocity. In particular, it justifies that the MAP estimate of the velocity is of form:

$$\boldsymbol{v}(\boldsymbol{r}) = \sum_{i=1}^{N} [\alpha_i \boldsymbol{G}_1^M (\boldsymbol{r} - \boldsymbol{r}_i) + \beta_i \boldsymbol{G}_2^M (\boldsymbol{r} - \boldsymbol{r}_i)]$$

as stated in equation (16), and the $\{\alpha\}, \{\beta\}$ are given by equations (17) for the dot, grating, and plaid stimuli. This generalizes the derivation of the original slow-and-smooth model [23, 24].

We first re-express the prior in terms of a differential operator (A.1), show that the MAP estimate can be expressed in terms of Green functions of the differential operator (A.2), and then show how to solve for the Green functions (A.3).

A.1 Re-express the Prior in terms of a Differential Operator $\mathcal{D}$ We use the quadratic form of the energy functions of the priors, see equations (7, 8, 9), to re-express them in terms of a differential operator $\mathcal{D}_{\mathcal{M}}$ indexed by the model M.

$$E(\{\boldsymbol{v}\}|M) = \int \boldsymbol{v} \cdot (\mathcal{D}_{\mathcal{M}} \boldsymbol{v}) d\boldsymbol{r}. \tag{25}$$

This re-expression is performed by functional integration (assuming suitable boundary conditions as $|\boldsymbol{r}| \mapsto \infty$) and yields:

$$\int |\boldsymbol{v}|^2 d\boldsymbol{r} = \int \boldsymbol{v}^T \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \boldsymbol{v} d\boldsymbol{r}$$

$$\int (|\nabla v_x|^2 + |\nabla v_y|^2) d\boldsymbol{r} = -\int \boldsymbol{v}^T \begin{pmatrix} \nabla^2 & 0 \\ 0 & \nabla^2 \end{pmatrix} \boldsymbol{v} d\boldsymbol{r}$$

$$\int 2 \frac{\partial v_x}{\partial y} \frac{\partial v_y}{\partial x} d\boldsymbol{r} = -\int \boldsymbol{v}^T \begin{pmatrix} 0 & \frac{\partial^2}{\partial x \partial y} \\ \frac{\partial^2}{\partial x \partial y} & 0 \end{pmatrix} \boldsymbol{v} d\boldsymbol{r} = \int 2 \frac{\partial v_x}{\partial x} \frac{\partial v_y}{\partial y} d\boldsymbol{r}$$

$$\int |\nabla^2 \boldsymbol{v}|^2 d\boldsymbol{r} = \int \boldsymbol{v}^T \begin{pmatrix} (\nabla^2)^2 & 0 \\ 0 & (\nabla^2)^2 \end{pmatrix} \boldsymbol{v} d\boldsymbol{r}$$

Combining the terms for the different priors $E(\boldsymbol{v}|M)$ gives the operators:

$$\mathcal{D}_t = \lambda \left(I - \mu \begin{pmatrix} \nabla^2 & 0 \\ 0 & \nabla^2 \end{pmatrix} + \eta \begin{pmatrix} (\nabla^2)^2 & 0 \\ 0 & (\nabla^2)^2 \end{pmatrix} \right)$$

$$\mathcal{D}_r = \lambda \left(I - \mu \begin{pmatrix} \nabla^2 & \frac{\partial^2}{\partial x \partial y} \\ \frac{\partial^2}{\partial x \partial y} & \nabla^2 \end{pmatrix} + \eta \begin{pmatrix} (\nabla^2)^2 & 0 \\ 0 & (\nabla^2)^2 \end{pmatrix} \right)$$

$$\mathcal{D}_e = \lambda \left(I - \mu \begin{pmatrix} \nabla^2 & -\frac{\partial^2}{\partial x \partial y} \\ -\frac{\partial^2}{\partial x \partial y} & \nabla^2 \end{pmatrix} + \eta \begin{pmatrix} (\nabla^2)^2 & 0 \\ 0 & (\nabla^2)^2 \end{pmatrix} \right)$$

A.2 The Euler-Lagrange Equations

The MAP estimate $\{\boldsymbol{v}\}$ for model M, $\{\boldsymbol{v}^*\} = \arg\max_{\{\boldsymbol{v}\}} P(\{\boldsymbol{v}\}|\{\boldsymbol{u}\}, M)$, is obtained by solving the Euler-Lagrange equations for the functional $E(\{\boldsymbol{u}\}|\{\boldsymbol{v}\}) + E(\{\boldsymbol{v}\}|M)$. This equation will depend on the form of the stimuli (e.g. dots and gratings).

For the dot stimuli, we express the functional in the form:

$$E(\{\boldsymbol{u}\}|\{\boldsymbol{v}\}) + E(\{\boldsymbol{v}\}|M) = \int \sum_{i=1}^{N} |\boldsymbol{v}(\boldsymbol{r}) - \boldsymbol{u}(\boldsymbol{r}_i)|^2 \delta(\boldsymbol{r} - \boldsymbol{r}_i) d\boldsymbol{r} + \int \boldsymbol{v} \cdot (\mathcal{D}_{\mathcal{M}} \boldsymbol{v}) d\boldsymbol{r}$$

(26)

This gives Euler-Lagrange equation:

$$\sum_{i=1}^{N} 2[\boldsymbol{v}(\boldsymbol{r}) - \boldsymbol{u}(\boldsymbol{r}_i)]\delta(\boldsymbol{r} - \boldsymbol{r}_i) + 2\mathcal{D}_{\mathcal{M}} \boldsymbol{v}(\boldsymbol{r}) = 0.$$

(27)

The solution can be obtained (extending [2] and [24]) by assuming a solution expressed as linear combination of vector valued Green's functions G_1^M, G_2^M:

$$\boldsymbol{v}(\boldsymbol{r}) = \sum_{i=1}^{N} [\alpha_i \boldsymbol{G}_1^M(\boldsymbol{r} - \boldsymbol{r}_i) + \beta_i \boldsymbol{G}_2^M(\boldsymbol{r} - \boldsymbol{r}_i)],$$

(28)

where $\{\alpha\}, \{\beta\}$ are coefficients and G_1^M, G_2^M are the solutions of

$$\mathcal{D}_M G_1(\boldsymbol{r}) = \begin{pmatrix} \delta(0) \\ 0 \end{pmatrix} = \boldsymbol{e}_1, \ \mathcal{D}_M G_2(\boldsymbol{r}) = \begin{pmatrix} 0 \\ \delta(0) \end{pmatrix} = \boldsymbol{e}_2$$

The form of the priors, and hence the Green functions, ensures that this solution is well-posed at the data points and falls off smoothly as $\boldsymbol{r} \mapsto \infty$. More specifically: (i) the second order derivative terms in the priors ensure that $G_1(0)$ and $G_2(0)$ are finite [24], and (ii) the slowness term ensures that the Green functions decrease to zero as $|\boldsymbol{r}| \mapsto \infty$ [24].

We obtain the linear equations for $\{\alpha\}, \{\beta\}$, given by equation (17), by substituting equation (28) into the Euler-Lagrange equations (27), using the Green functions from equation (9).

For the grating stimulus, we obtain similar results. The energy functional becomes:

$$E(\{\boldsymbol{u}\}|\{\boldsymbol{v}\}) + E(\{\boldsymbol{v}\}|M) = \int \sum_{i=1}^{N} |\boldsymbol{v}(\boldsymbol{r}) \cdot \hat{\boldsymbol{u}}(\boldsymbol{r}_i) - u(\boldsymbol{r}_i)|^2 \delta(\boldsymbol{r} - \boldsymbol{r}_i) d\boldsymbol{r} + \int \boldsymbol{v} \cdot (\mathcal{D}_{\mathcal{M}} \boldsymbol{v}) d\boldsymbol{r}.$$

(29)

This gives Euler-Lagrange equations:

$$\sum_{i=1}^{N} \{[\boldsymbol{v}(\boldsymbol{r}) \cdot \hat{\boldsymbol{u}}(\boldsymbol{r}_i)]\hat{\boldsymbol{u}}(\boldsymbol{r}_i) - u(\boldsymbol{r}_i)\}\delta(\boldsymbol{r} - \boldsymbol{r}_i) + \mathcal{D}_M \boldsymbol{v}(\boldsymbol{r}) = 0.$$

(30)

The solution $\boldsymbol{v}$ can be expressed in form given by equation (28) with the same Green functions from equation (9). Substituting the form of $\boldsymbol{v}$ into the Euler-Lagrange equation (30) gives the linear equation (19) for the $\{\alpha\}, \{\beta\}$.

A.3. Solving for the Green Functions We solve equation (9) for the Green functions by Fourier transforms. Denote $G_1(\boldsymbol{r}) = (G_{1x}(\boldsymbol{r}), \ G_{1y}(\boldsymbol{r}))^T$ (we drop the model index ^{M} to simplify notation). This gives the following solutions for the three different cases:

1. *Slow-and-smooth-translation*: We apply Fourier transforms to the equation for the Green functions:

$$\lambda \begin{pmatrix} 1 - \mu\nabla^2 + \eta(\nabla^2)^2 & 0 \\ 0 & 1 - \mu\nabla^2 + \eta(\nabla^2)^2 \end{pmatrix} \begin{pmatrix} G_{1x} \\ G_{1y} \end{pmatrix} = \begin{pmatrix} \delta(0) \\ 0 \end{pmatrix}$$

    to obtain linear equations for the Fourier transforms $\mathcal{F}G(\omega_x, \omega_y)$ (where $\omega^2 = \omega_x^2 + \omega_y^2$):

$$\lambda \begin{pmatrix} 1 + \mu\omega^2 + \eta\omega^4 & 0 \\ 0 & 1 + \mu\omega^2 + \eta\omega^4 \end{pmatrix} \begin{pmatrix} \mathcal{F}G_{1x} \\ \mathcal{F}G_{1y} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

    which is solved to give

$$\begin{pmatrix} \mathcal{F}G_{1x} \\ \mathcal{F}G_{1y} \end{pmatrix} = \frac{1}{\lambda} \begin{pmatrix} \frac{1}{1+\mu\omega^2+\eta\omega^4} \\ 0 \end{pmatrix}$$

    with similar solution for $\mathcal{F}G_2$.

2. *Slow-and-smooth-rotation*: we re-express the Green function equation in fourier space

$$\lambda \begin{pmatrix} 1 - \mu\nabla^2 + \eta(\nabla^2)^2 & -\mu\frac{\partial^2}{\partial x \partial y} \\ -\mu\frac{\partial^2}{\partial x \partial y} & 1 - \mu\nabla^2 + \eta(\nabla^2)^2 \end{pmatrix} \begin{pmatrix} G_{1x} \\ G_{1y} \end{pmatrix} = \begin{pmatrix} \delta(0) \\ 0 \end{pmatrix}$$

$$\lambda \begin{pmatrix} 1 + \mu\omega^2 + \eta\omega^4 & \mu\omega_x\omega_y \\ \mu\omega_x\omega_y & 1 + \mu\omega^2 + \eta\omega^4 \end{pmatrix} \begin{pmatrix} \mathcal{F}G_{1x} \\ \mathcal{F}G_{1y} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

    defining

$$d_r = \det\left[\begin{pmatrix} 1 + \mu\omega^2 + \eta\omega^4 & -\mu\omega_x\omega_y \\ -\mu\omega_x\omega_y & 1 + \mu\omega^2 + \eta\omega^4 \end{pmatrix} \right]$$

    gives the solution

$$\begin{pmatrix} \mathcal{F}G_{1x} \\ \mathcal{F}G_{1y} \end{pmatrix} = \frac{1}{\lambda d_r} \begin{pmatrix} 1 + \mu\omega^2 + \eta\omega^4 & -\mu\omega_x\omega_y \\ -\mu\omega_x\omega_y & 1 + \mu\omega^2 + \eta\omega^4 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

3. *Slow-and-smooth-expansion*: this is a simple modification of the rotation case:

$$\lambda \begin{pmatrix} 1 + \mu\nabla^2 + \eta(\nabla^2)^2 & -\mu\frac{\partial^2}{\partial x \partial y} \\ -\mu\frac{\partial^2}{\partial x \partial y} & 1 + \mu\nabla^2 + \eta(\nabla^2)^2 \end{pmatrix} \begin{pmatrix} \mathcal{F}G_{1x} \\ \mathcal{F}G_{1y} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

yields the solution

$$\begin{pmatrix} \mathcal{F}G_{1x} \\ \mathcal{F}G_{1y} \end{pmatrix} = \frac{1}{\lambda d_e} \begin{pmatrix} 1 + \mu\omega^2 + \eta\omega^4 & \mu\omega_x\omega_y \\ \mu\omega_x\omega_y & 1 + \mu\omega^2 + \eta\omega^4 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

Note that the following equalities hold for both rotation and expansion:

$$G_{1x}(x, y) = G_{2y}(y, x), \; G_{1y}(x, y) = G_{2x}(x, y)$$

B. Model Selection

The model evidence is

$$p(\{\boldsymbol{u}\}) = \int p(\{\boldsymbol{u}\}|\{\boldsymbol{v}\})p(\{\boldsymbol{v}\})d\{\boldsymbol{v}\}$$

which we re-express in term of (A, B)

$$p(\{\boldsymbol{u}\}) = \int p(\{\boldsymbol{u}\}|A, B)p(A, B)dAdB$$

in which

$$p(\{\boldsymbol{u}\}|A, B) = \frac{1}{Z_1}\exp(-E(\{\boldsymbol{u}\}|A, B)/T), \; p(A, B) = \frac{1}{Z_2}\exp(-E(A, B)/T)$$

where, for the random dot stimuli,

$$E(\{\boldsymbol{u}\}|A, B) = \sum_{i=1}^{N}|\boldsymbol{v}(\boldsymbol{r}_i) - \boldsymbol{u}(\boldsymbol{r}_i)|^2 = \Phi^T G^2 \Phi - 2U^T G\Phi + U^T U, \qquad (31)$$

where $\Phi = (AB)^T$.

We can re-express this as:

$$E(A, B) = \Phi^T G\Phi,$$

using $\boldsymbol{v}(\boldsymbol{r}) = \sum_{i=1}^{N}[\alpha_i G_1(\boldsymbol{r} - \boldsymbol{r}_i) + \beta_i G_2(\boldsymbol{r} - \boldsymbol{r}_i)]$ and $\mathcal{D}\boldsymbol{v}(\boldsymbol{r}) = \sum_{i=1}^{N}[\alpha_i \delta(\boldsymbol{r} - \boldsymbol{r}_i)e_1 + \beta_i \delta(\boldsymbol{r} - \boldsymbol{r}_i)e_2]$.

The partition functions are of form:

$$Z_1 = \int \exp(-E(\{u\}|A, B)/T)d\{u\} = \int \exp(-(U - V)^T(U - V)/T)dU = (\pi T)^N$$

$$Z_2 = \int \exp(-E(A, B)/T)dAdB = \int \exp(-\Phi^T G\Phi/T)d\Phi = \frac{(\pi T)^N}{\sqrt{\det(G)}}$$

We now re-express the model evidence $p(\{\boldsymbol{u}\})$,

$$p(\{\boldsymbol{u}\}) = \int \frac{1}{Z_1}\exp(-E(\{\boldsymbol{u}\}|A, B)/T)\frac{1}{Z_2}\exp(-E(A, B)/T)dAdB$$

$$= \frac{1}{Z_1 Z_2}\int \exp[-\frac{1}{T}(\Phi^T(G^2 + G)\Phi - 2U^T G\Phi + U^T U)]d\Phi$$

denoting

$$\Sigma = G^2 + G = GH, \ H = G + I, \ \Theta = \Sigma^{-1} GU = H^{-1} U$$

gives

$$
\begin{aligned}
p(\{\boldsymbol{u}\}) &= \frac{1}{Z_1 Z_2} \int \exp[-\frac{1}{T}(\Phi^T \Sigma \Phi - 2U^T G \Phi + U^T U)] d\Phi \\
&= \frac{1}{Z_1 Z_2} \int \exp[-\frac{1}{T}((\Phi - \Theta)^T \Sigma (\Phi - \Theta) + U^T U - \Theta^T \Sigma \Theta)] d\Phi \\
&= \frac{1}{Z_1 Z_2} \exp[-\frac{1}{T}(U^T U - \Theta^T \Sigma \Theta)] \int \exp\{-\frac{1}{T}[(\Phi - \Theta)^T \Sigma (\Phi - \Theta)]\} d\Phi \\
&= \frac{1}{(\pi T)^N} \frac{\sqrt{\det(G)}}{(\pi T)^N} \exp[-\frac{1}{T}(U^T U - \Theta^T \Sigma \Theta)] \frac{(\pi T)^N}{\sqrt{\det(\Sigma)}} \\
&= \frac{1}{(\pi T)^N \sqrt{\det(G + I)}} \exp[-\frac{1}{T}(U^T U - U^T H^{-1} U)].
\end{aligned}
$$

Modification for the Grating Stimulus In the model selection stage we modify the energy to be,

$$
\begin{aligned}
E(\{\boldsymbol{u}\}|A, B) &= \sum_{i=1}^{N} |\boldsymbol{v}(\boldsymbol{r}_i) \cdot \hat{\boldsymbol{u}}(\boldsymbol{r}_i) - u(\boldsymbol{r}_i)|^2 \\
&= \Phi^T \tilde{G}^T \tilde{G} \Phi - 2U^T \tilde{G} \Phi + U^T U
\end{aligned}
$$

$$E(A, B) = \int \boldsymbol{v} \cdot (\mathcal{D}\boldsymbol{v}) d\boldsymbol{r} = \Phi^T G \Phi$$

The partition functions Z_1 and Z_2 remain the same. So

$$
\begin{aligned}
p(\{\boldsymbol{u}\}) &= \int \frac{1}{Z_1} \exp(-E(\{\boldsymbol{u}\}|A, B)/T) \frac{1}{Z_2} \exp(-E(A, B)/T) dA dB \\
&= \frac{1}{Z_1 Z_2} \int \exp[-\frac{1}{T}(\Phi^T(\tilde{G}^T \tilde{G} + G)\Phi - 2U^T \tilde{G} \Phi + U^T U)] d\Phi
\end{aligned}
$$

denoting

$$\tilde{\Sigma} = \tilde{G}^T \tilde{G} + G, \ \tilde{\Theta} = \tilde{\Sigma}^{-1} \tilde{G}^T U$$

and (noting that $\tilde{\Sigma}$ is symmetric even though $\tilde{G}$ is not) we get:

$$
\begin{aligned}
p(\{u\}) &= \frac{1}{Z_1 Z_2} \int \exp[-\frac{1}{T}(\Phi^T \tilde{\Sigma} \Phi - 2U^T \tilde{G} \Phi + U^T U)] d\Phi \\
&= \frac{1}{Z_1 Z_2} \int \exp[-\frac{1}{T}((\Phi - \tilde{\Theta})^T \tilde{\Sigma}(\Phi - \tilde{\Theta}) + U^T U - \tilde{\Theta}^T \tilde{\Sigma} \tilde{\Theta})] d\Phi \\
&= \frac{1}{Z_1 Z_2} \exp[-\frac{1}{T}(U^T U - \tilde{\Theta}^T \tilde{\Sigma} \tilde{\Theta})] \int \exp\{-\frac{1}{T}[(\Phi - \tilde{\Theta})^T \tilde{\Sigma}(\Phi - \tilde{\Theta})]\} d\Phi \\
&= \frac{1}{(\pi T)^N} \frac{\sqrt{\det(G)}}{(\pi T)^N} \exp[-\frac{1}{T}(U^T U - \tilde{\Theta}^T \tilde{\Sigma} \tilde{\Theta})] \frac{(\pi T)^N}{\sqrt{\det(\tilde{\Sigma})}} \\
&= \frac{1}{(\pi T)^N} \sqrt{\det(\frac{G}{\tilde{\Sigma}})} \exp[-\frac{1}{T}(U^T U - U^T \tilde{G} \tilde{\Sigma}^{-1} \tilde{G}^T U)]
\end{aligned}
$$

Appendix C: Least-Square Fit of Rotation/Expansion

This appendix answers the question of how to estimate the coefficients of the expansion and rotation motions from the estimated velocity fields. More precisely, we need to estimate the centers (x_c, y_c) of expansion/rotation and the expansion/rotation coefficients e, ω. We perform this by least squares fitting.

C1. Rotation

We need to estimate the center (x_c, y_c) and rotation ω from a velocity field. We define

$$
\boldsymbol{u}(x, y) = (\omega(y_c - y), \ \omega(x - x_c))
$$

and an energy function:

$$
E(x_x, y_c, \omega) = \sum_{i=1}^{N} \{[v_x(\boldsymbol{r}_i) - \omega(y_c - y_i)]^2 + [v_y(\boldsymbol{r}_i) - \omega(x_i - x_c)]^2\}
$$

We estimate x_c, y_c and ω by minimizing $E(.,.,.)$. This is performed analytically by solving the equations $\frac{\partial E}{\partial x_c} = \frac{\partial E}{\partial y_c} = \frac{\partial E}{\partial \omega} = 0$. The derivatives with respect to (x_c, y_c) can be computed and set to zero yielding the equations:

$$
\frac{\partial E}{\partial x_c} = 2 \sum_{i=1}^{N} [v_y(\boldsymbol{r}_i) - \omega(x_i - x_c)] \cdot \omega = 2\omega N(\bar{v}_y - \omega \bar{x} + \omega x_c) = 0
$$

$$
\frac{\partial E}{\partial y_c} = 2 \sum_{i=1}^{N} [v_x(\boldsymbol{r}_i) - \omega(y_c - y_i)] \cdot (-\omega) = -2\omega N(\bar{v}_x - \omega y_c + \omega \bar{y}) = 0
$$

This gives the solutions (x_c, y_c) as functions of ω and the means $\bar{x} = 1/N \sum_{i=1}^{N} x_i$, $\bar{y} = 1/N \sum_{i=1}^{N} y_i, \bar{v}_x = 1/N \sum_{i=1}^{N} v_x(\boldsymbol{r}_i), \bar{v}_y = 1/N \sum_{i=1}^{N} v_y(\boldsymbol{r}_i)$ of the data points and their estimated velocities. The solutions are given by:

$$\begin{cases} x_c & = \bar{x} - \frac{\bar{v}_y}{\omega} \\ y_c & = \bar{y} + \frac{\bar{v}_x}{\omega} \end{cases}$$

To solve for ω, we calculate $\frac{\partial E}{\partial \omega}$ and set it to zero. This gives

$$\frac{\partial E}{\partial \omega} = -2 \sum_{i=1}^{N} \{[v_x(\boldsymbol{r}_i) - \omega(y_c - y_i)](y_c - y_i) + [v_y(\boldsymbol{r}_i) - \omega(x_i - x_c)](x_i - x_c)\}$$

$$= -2N[y_c\bar{v}_x - x_c\bar{v}_y - \overline{yv_x} + \overline{xv_y} - \omega(x_c^2 + y_c^2) + 2\omega(x_c\bar{x} + y_c\bar{y}) - \omega(\overline{x^2} + \overline{y^2})] = 0$$

Plugging in x_c and y_c yields

$$0 = (\bar{y}\bar{v}_x - \overline{yv_x}) - (\bar{x}\bar{v}_y - \overline{xv_y}) + \omega(\bar{x}^2 + \bar{y}^2) - \omega(\overline{x^2} + \overline{y^2}) \qquad (32)$$

hence

$$\omega = \frac{(\overline{xv_y} - \bar{x}\bar{v}_y) - (\overline{yv_x} - \bar{y}\bar{v}_x)}{(\overline{x^2} + \overline{y^2}) - (\bar{x}^2 + \bar{y}^2)}$$

C2. Expansion

Similarly to the rotation case, define an expansion velocity field with three parameters x_c, y_c and k.

$$\boldsymbol{u}(x, y) = (k(x - x_c), \; k(y - y_c))$$

Define an energy

$$E(x_c, y_c, k) = \sum_{i=1}^{N} \{[v_x(\boldsymbol{r}_i) - k(x_i - x_c)]^2 + [v_y(\boldsymbol{r}_i) - k(y_i - y_c)]\}^2$$

The x_c, y_c and k that minimize E satisfy

$$\frac{\partial E}{\partial x_c} = \frac{\partial E}{\partial y_c} = \frac{\partial E}{\partial k} = 0$$

As before, we solve for (x_c, y_c) as functions of k and the means of the data points and their velocities.

$$\frac{\partial E}{\partial x_c} = 2 \sum_{i=1}^{N} [v_x(\boldsymbol{r}_i) - k(x_i - x_c)] \cdot k = 2kN(\bar{v}_x - k\bar{x} + kx_c) = 0$$

$$\frac{\partial E}{\partial y_c} = 2 \sum_{i=1}^{N} [v_y(\boldsymbol{r}_i) - k(y_i - y_c)] \cdot k = 2kN(\bar{v}_y - k\bar{y} + ky_c) = 0 \qquad (33)$$

This leads to

$$\begin{cases} x_c & = \bar{x} - \frac{\bar{v}_x}{k} \\ y_c & = \bar{y} - \frac{\bar{v}_y}{k} \end{cases}$$

Then we solve for k by setting

$$\frac{\partial E}{\partial k} = -2 \sum_{i=1}^{N} \{[v_x(\boldsymbol{r}_i) - k(x_i - x_c)](x_i - x_c) + [v_y(\boldsymbol{r}_i) - k(y_i - y_c)](y_i - y_c)\}$$

$$= -2N[\overline{xv_x} - x_c\bar{v}_x + \overline{yv_y} - y_c\bar{v}_y - k(\bar{x^2} + \bar{y^2}) + 2k(x_c\bar{x} + y_c\bar{y}) - k(x_c^2 + y_c^2)] = 0$$

and plugging in x_c and y_c gives

$$0 = \overline{xv_x} + \overline{yv_y} - (\bar{x} - \frac{\bar{v}_x}{k})\bar{v}_x - (\bar{y} - \frac{\bar{v}_y}{k})\bar{v}_y - k(\bar{x^2} + \bar{y^2})$$

$$+ 2k[(\bar{x} - \frac{\bar{v}_x}{k})\bar{x} + (\bar{y} - \frac{\bar{v}_y}{k})\bar{y}] - k[(\bar{x} - \frac{\bar{v}_x}{k})^2 + (\bar{y} - \frac{\bar{v}_y}{k})^2]$$

hence

$$k = \frac{(\overline{xv_x} - \bar{x}\bar{v}_x) + (\overline{yv_y} - \bar{y}\bar{v}_y)}{(\bar{x^2} + \bar{y^2}) - (\bar{x}^2 + \bar{y}^2)}$$

References

1. J.F. Barraza and N.M. Grzywacz. Measurement of angular velocity in the perception of rotation. Vision Research, 42.2002.
2. J. Duchon. Lecture Notes in Math. 571, (eds Schempp, W. and Zeller, K.) 85-100. Springer-Verlag, Berlin, 1977.
3. C. J. Duffy, and R. H. Wurtz. Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selectivity to large field stimuli. Journal of Neurophysiology. 65, 1329-1345. 1991.
4. T. Freeman, and M. Harris. Human sensitivity to expanding and rotating motion: effect of complementary masking and directional structure. Vision research, 32, 1992.
5. E.C. Hildreth. Computations Underlying the Measurement of Visual Motion. Artificial Intelligence. 23(3). pp 309-354. 1984.
6. B. Horn and B. Schunck. Determining Optical Flow. Artificial Intelligence 17. 1981.
7. D. Knill and W. Richards (Eds). Perception as Bayesian Inference. Cambridge University Press, 1996.
8. A. Lee, A. Yuille, and H. Lu. Superior perception of circular/radial than translational motion cannot be explained by generic priors. VSS 2008.
9. H. Lu and A.L. Yuille. Ideal Observers for Detecting Motion: Correspondence Noise. NIPS 2005.
10. M. C. Morrone, D. C. Burr, and L. Vaina. Two stages of visual processing for radial and circular motion. Nature, 376, 507-509. 1995.
11. M. Morrone, M. Tosetti, D. Montanaro, A. Fiorentini, G. Cioni, and D. C. Burr. A cortical area that responds specifically to optic flow revealed by fMRI. Nature Neuroscience, 3, 1322 -1328. 2000.
12. S. Nishida, K. Amano, M. Edwards, and D.R. Badcock. Global motion with multiple Gabors - A tool to investigate motion integration across orientation and space. VSS 2006.

13. V.S. Ramachandran and S.M. Anstis. The perception of apparent motion. Scientific American, 254, 102-109. 1986.
14. C. E. Rasmussen and C.K.I. Williams. Gaussian Processes for Machine Learning. The MIT Press, 2006.
15. S. Roth and M.J. Black. On the Spatial Statistics of Optical Flow. International Conference of Computer Vision. 2005.
16. R. Sekuler, S.N.J. Watamaniuk and R. Blake. Perception of Visual Motion. In Steven's Handbook of Experimental Psychology. Third edition. H. Pashler, series editor. S. Yantis, volume editor. J. Wiley Publishers. New York. 2002.
17. A.A. Stocker and E.P. Simoncelli. Noise characteristics and prior expectations in human visual speed perception Nature Neuroscience, vol. 9(4), pp. 578–585, Apr 2006.
18. A.A. Stocker, and E. Simoncelli. A Bayesian model of conditioned perception. *Proceedings of Neural Information Processing Systems*. 2007.
19. K. Tanaka, Y. Fukada, and H. Saito. Underlying mechanisms of the response specificity of expansion/contraction and rotation cells in the dorsal part of the MST area of the macaque monkey. Journal of Neurophysiology. 62, 642-656. 1989.
20. S. Ullman. The Interpretation of Structure from Motion. PhD Thesis. MIT. 1979.
21. Y. Weiss, and E.H. Adelson. Slow and smooth: A Bayesian theory for the combination of local motion signals in human vision Technical Report 1624. Massachusetts Institute of Technology. 1998.
22. Y. Weiss, E.P. Simoncelli, and E.H. Adelson. Motion illusions as optimal percepts. Nature Neuroscience, 5, 598-604. 2002.
23. A.L. Yuille and N.M. Grzywacz. A computational theory for the perception of coherent visual motion. *Nature*, 333,71-74. 1988.
24. A.L. Yuille and N.M. Grzywacz. A Mathematical Analysis of the Motion Coherence Theory. International Journal of Computer Vision. 3. pp 155-175. 1989.
25. A.L. Yuille and S. Ullman. Rigidity and Smoothness of Motion: Justifying the Amoothness Assumption in Motion Measurement. (Chp 8). Image Understanding 1989. Ed. S. Ullman and W. Richards. 1989.